

## University of Groningen

### Running Repairs

Healey, Patrick G. T.; Mills, Gregory J.; Eshghi, Arash; Howes, Christine

*Published in:*  
Topics in Cognitive Science

*DOI:*  
[10.1111/tops.12336](https://doi.org/10.1111/tops.12336)

**IMPORTANT NOTE:** You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2018

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Healey, P. G. T., Mills, G. J., Eshghi, A., & Howes, C. (2018). Running Repairs: Coordinating Meaning in Dialogue. *Topics in Cognitive Science*, 10(2), 367-388. <https://doi.org/10.1111/tops.12336>

#### Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

#### Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*



Topics in Cognitive Science 10 (2018) 367–388

Copyright © 2018 The Authors. Topics in Cognitive Science published by Wiley Periodicals, Inc. on behalf of Cognitive Science Society

ISSN:1756-8757 print / 1756-8765 online

DOI: 10.1111/tops.12336

This article is part of the topic “Miscommunication,” Patrick Healey, Jan De Ruiter and Gregory Mills (Topic Editors). For a full listing of topic papers, see <https://onlinelibrary.wiley.com/toc/17568765/2018/10/2>

## Running Repairs: Coordinating Meaning in Dialogue

Patrick G. T. Healey,<sup>a</sup> Gregory J. Mills,<sup>b</sup> Arash Eshghi,<sup>c</sup> Christine Howes<sup>d</sup>

<sup>a</sup>*School of Electronic Engineering and Computer Science, Queen Mary University of London*

<sup>b</sup>*Language Technology — Computational Linguistics, University of Groningen*

<sup>c</sup>*Computer Science, Heriot-Watt University*

<sup>d</sup>*General Linguistics, University of Gothenburg*

Received 18 March 2016; received in revised form 19 January 2018; accepted 14 February 2018

---

### Abstract

People give feedback in conversation: both positive signals of understanding, such as nods, and negative signals of misunderstanding, such as frowns. How do signals of understanding and misunderstanding affect the coordination of language use in conversation? Using a chat tool and a maze-based reference task, we test two experimental manipulations that selectively interfere with feedback in live conversation: (a) “Attenuation” that replaces positive signals of understanding such as “right” or “okay” with weaker, more provisional signals such as “errr” or “umm” and (2) “Amplification” that replaces relatively specific signals of misunderstanding from clarification requests such as “on the left?” with generic signals of trouble such as “huh?” or “eh?”. The results show that Amplification promotes rapid convergence on more systematic, abstract ways of describing maze locations while Attenuation has no significant effect. We interpret this as evidence that “running repairs”—the processes of dealing with misunderstandings on the fly—are key drivers of semantic coordination in dialogue. This suggests a new direction for experimental work on conversation and a productive way to connect the empirical accounts of Conversation Analysis with the representational and processing concerns of Formal Semantics and Psycholinguistics.

**Keywords:** Dialogue; Miscommunication; Repair

---

---

Correspondence should be sent to Patrick G. T. Healey, School of Electronic Engineering and Computer Science, Queen Mary University of London, Mile End Road, London E1 4NS. E-mail: [p.healey@qmul.ac.uk](mailto:p.healey@qmul.ac.uk)

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

## 1. Introduction

People do not use the same words in the same way. Differences in life history, cultural background, individual physiology, and social communities can all contribute to differences in people's language use (Clark, 1998). This apparently simple observation has significant consequences. It raises foundational questions about what it means to say that two people speak the same language or mean the same thing. It also raises practical questions. How can people communicate if there really are widespread individual differences in language use? This paper engages with these issues through an experimental investigation of how different kinds of feedback contribute to the coordination of language use in conversation. More specifically, we investigate what signals of understanding and misunderstanding contribute to people's ability to coordinate their descriptions of locations in a simple schematic maze (Anderson & Garrod, 1987).

The results provide evidence that natural signals of understanding and misunderstanding have systematically different effects on the kind of language people use to describe where they are in the maze. Signals of misunderstanding, in particular, appear to have a special role in coordinating changes to the semantics of maze location descriptions. We interpret this finding as providing support for a general *Running Repairs Hypothesis*: coordination of language use depends primarily on processes used to deal with misunderstanding on the fly and only secondarily on those associated with signaling understanding. The basic intuition behind this hypothesis is that differences between different people's patterns of language use are more efficiently dealt with by identifying and dealing with specific local points of difference (i.e., misunderstandings) than by confirming apparent points of commonality (i.e., understanding). Public signals of misunderstanding can provide joint constraints that are useful for driving a progressive narrowing down of possible interpretations. Public signals of understanding, by contrast, primarily provide confirmation of existing interpretations rather than specific momentum for change and are probably most useful after people have converged by other means.

The Running Repairs Hypothesis is directly inspired by the Conversation Analytic (CA) account of *repair*: the procedures whereby people detect and deal with difficulties with mutual understanding (Sacks, Schegloff, & Jefferson, 1974a; Schegloff, 1987, 1995). It is an explicit attempt to connect CA insights about repair with the representational and processing concerns of the cognitive sciences. The potential interest of this connection for cognitive scientists depends partly on whether formal and computational models can be developed that provide convincing repair-driven accounts of semantic coordination (see Ginzburg, 2012; Eshghi, Howes, Hough, Gregoromichelaki, & Purver, 2015; Howes & Eshghi, 2017; and also Ginzburg & Kolliaku, 2018; Larson, 2018, this volume). It also depends on providing experimental evidence of causal connections between repair processes and specific patterns of language use in conversation.

## 2. Coordinating language use

It is useful to distinguish two broad psycholinguistic approaches to conversation that we gloss here as *Aggregate* and *Interactional*. This is not an individual vs. social distinction; both approaches aim to explain coordinated language processing in conversation. It is also unlikely that any psycholinguistic model of conversation falls exclusively into only one of these categories; the distinction is introduced primarily to help clarify what is at issue.

Aggregate approaches emphasize the use of established models of individual language processing as a way of accounting for dialogue. They use the same basic cognitive mechanisms that explain individual lexical, syntactic, and semantic processing and then model dialogue as aggregations of those mechanisms acting in concert (Barr & Keysar, 2002; Christiansen & Chater, 2016; Horton & Gerrig, 2005; Pickering & Garrod, 2004, 2006). The key advantage of this general approach is simplicity as it builds on relatively well-understood mechanisms that have been extensively tested experimentally. Aggregate models have been used, among other things, to explain patterns of language coordination in the Maze Task (see discussion below and also, for example, Garrod & Doherty, 1994; Pickering & Garrod, 2004).

In contrast to this, Interactional approaches propose that dialogue presents processing challenges that are qualitatively different from individual language processing and require different mechanisms to explain them (Bavelas, 2007; Clark & Schaefer, 1989; Eshghi & Healey, 2016; Goodwin, 1979; Healey, 2008; Schober & Brennan, 2003).

The processes most commonly emphasized in Interactional accounts are the forms of interactive feedback that people provide to each other during conversation. For example, addressees normally provide ongoing evidence of how well they are following a speaker's turn by producing concurrent, carefully timed, and contextually appropriate signals of understanding using gaze, facial expressions, and *backchannels* such as nods, "yeah"s, and "aha"s (e.g., Bavelas et al., 2000; Wilkes-Gibbs & Clark, 1992; Yngve, 1970). Alternatively, if addressees encounter problems in interpreting a speaker's ongoing turn, they typically produce concurrent signals of misunderstanding such as raised eyebrows, puzzled looks, or overt clarification questions such as "eh?" and "sorry?" (Dingemanse, Torreira, & Enfield, 2013; Drew, 1997; Purver, Ginzburg & Healey, 2003).

This kind of structured conversational feedback is characteristic of natural conversation in almost any context (Bavelas et al., 2000; Clark, 1996; Colman & Healey, 2011; Dingemanse et al., 2013; Kendrick, 2015; Schegloff, 1993) but largely absent from language in other contexts of use, such as speeches or narration, and rarely encountered in standard psycholinguistic laboratory tasks.

By default Aggregate accounts idealize to cases in which processes of language production and comprehension do not diverge in significant ways between different people; a consequence of the claim that both inter-individual processing and intra-individual processing can be modeled using the same lexical, syntactic, semantic, and contextual representations, and the same processing mechanisms (Pickering & Garrod, 2004, 2006, 2013).

However, this makes accounting for structured conversational feedback phenomena problematic.

If this idealization is satisfied, it is unclear why people should provide frequent ongoing evidence of understanding in conversation. By hypothesis, intra- and inter-person language processing use the same mechanisms, so differences in the form or frequency of feedback have to be explained as somehow epiphenomenal or auxiliary to normal language processing. Where this idealization is not satisfied because of misunderstandings due to mismatches in language use, the primary Aggregate mechanisms of coordination are poorly equipped to deal with them (e.g., automatic mutual priming in Pickering and Garrod [2004, 2006] or rapid “efferent copy” predictions that shadow production and comprehension in Pickering and Garrod [2013]). Instead, alternative Interactional processes for detecting and repairing problems are needed and are often cited as auxiliary mechanisms.

In contrast to this, the Running Repairs Hypothesis treats interactive feedback as central to language processing in conversation, a position closely associated with Clark’s grounding model of dialogue (Clark, 1996). It further proposes that the specific forms of interactive feedback associated with misunderstanding have a special status in explaining the form of language coordination that emerges in conversation; that is, that misunderstandings are the primary drivers of changes in the form of semantic coordination that develops through interaction (Healey, 1997, 2008; Mills, 2013).

We explore these claims through experiments on referring expressions in the Maze Task. These experiments provide a direct test of the causal effects of feedback on coordination by selectively interfering with these mechanisms during live text–chat conversations. They also provide a direct contrast of the contribution of positive evidence of understanding such as “yeah” and “aha” with negative evidence of misunderstanding such as “wot?” and “sorry?” to language coordination in conversation.

### **3. Language coordination experiments**

The most common paradigm in psycholinguistic experiments on dialogue involves investigation of how people coordinate their use of referring expressions. In these tasks, people try to describe something—typically an object, figure, or location—in a way that allows their conversational partner to pick the same thing out from a set of alternatives (Clark & Wilkes-Gibbs, 1986; Garrod & Anderson, 1987; Horton & Gerrig, 2005; Krauss & Weinheimer, 1966; Metzing & Brennan, 2003).

The recurrent finding in these experiments is that the form of referring expressions people produce systematically changes over time. What processes drive this change? Aggregate accounts argue that they are best explained in terms of mechanisms such as recent activation of a word or syntactic structure, general linguistic precedent, or memory constraints (Barr & Keysar, 2002; Horton & Gerrig, 2005; Pickering & Garrod, 2004). Interactional accounts claim that structured feedback processes, and specific histories of their use in specific interactions, play an essential role (Clark & Wilkes-Gibbs, 1986; Metzing & Brennan, 2003; Schober & Clark, 1989).

The best known Interactional account of these processes is the collaborative model of grounding (Clark, 1996; Clark & Schaefer, 1989; Schober & Clark, 1989; Clark & Wilkes-Gibbs, 1986). The collaborative model emphasizes the role of positive feedback. In order for a speaker's referring expression to be accepted as common usage in a conversation, addressees must provide appropriate evidence that it has been heard and understood. On the assumption that people seek to minimize the joint effort needed to communicate, this provides an account of how referring expressions can become highly abbreviated over repeated references to an object. Importantly, this pattern of change in referring expressions is sensitive to the provision of appropriate feedback from specific conversational partners (Schober & Clark, 1989; Clark & Wilkes-Gibbs, 1986; Eshghi & Healey, 2016).

When a referring expression fails, positive evidence, or its absence, is of limited use because it provides few constraints on how people should revise what they said to help their partner understand it. Without any other information, a speaker is left with trying successive alternative referring expressions until they find something that works (c.f. Lewis's [1969] analysis of coordination of linguistic conventions). In effect, coordination is modeled as a problem of an individual selecting from an existing repertoire of ways of referring to things or inventing a new referring expression in the hope that they can unilaterally identify a scheme that also works for collaboration with their partner.

The Running Repairs Hypothesis proposes that the bilateral processes of signaling and resolving misunderstandings are key to people's ability to build new conventions or "sub-languages" on-the-fly in task-oriented dialogues (Healey, 1997, 2008; Healey, Swoboda, Umata, & King, 2007; Mills, 2013). Negative evidence provides people with specific new information about differences in language use which provide a more fine-grained, cross-speaker basis for revising their referring expressions than positive evidence can. We apply this argument in the context of coordination of language use in the Maze task.

#### **4. The Maze Task**

The Maze Task was devised by Garrod and Anderson (Anderson & Garrod, 1987; Garrod & Anderson, 1987) as an interactive game in which pairs of people collaborate to navigate through a maze to reach a goal point (Fig. 1). They cannot see each other's positions in the maze and in order to reach their respective goals they must describe to each other where switch points are that open and close gates in each other's mazes. As a result, they repeatedly exchange location descriptions. Importantly the configuration of the Maze changes over trials and this pushes participants toward developing systematic approaches to describing sequences of different locations in sequences of different mazes.

The referring expressions people produce to describe locations in this task can be reliably classified into four general classes; Figural, Path, Line, and Matrix (Anderson & Garrod, 1987; Garrod & Anderson, 1987). They are coded into these categories according to the surface form of the location descriptions, including choice of words (e.g. "box" vs. "row"), use of counting systems (none vs. ordinal vs. cardinal), and the procedure they use to locate a particular position (landmark based, path based, line based, or coordinate

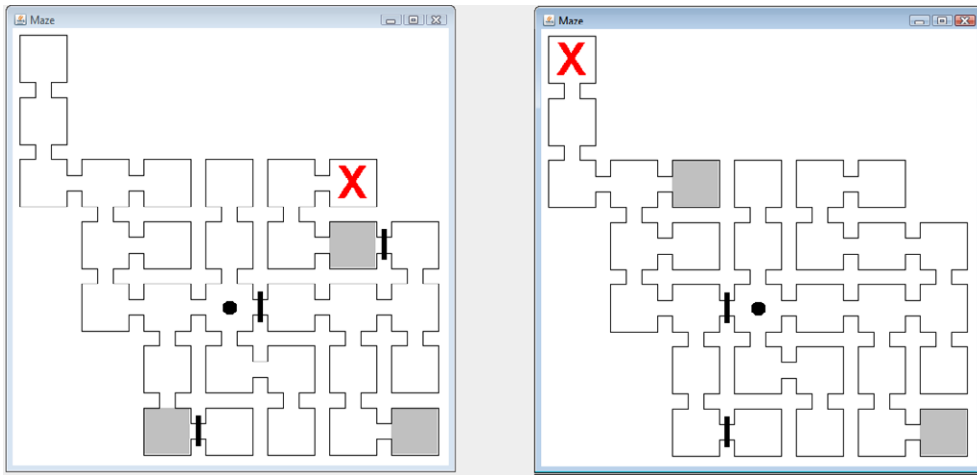


Fig. 1. Pair of maze configurations. vertical black bars = gates, gray squares = switch points.

based). We illustrate each description type with examples drawn from the text chat experimental corpus described below.

Figural and Path descriptions both rely on the specific configuration of boxes and links in each instance of the maze. Figural descriptions make use of salient landmarks or shapes in a maze as a way to identify target locations as in or near a landmark, for example, “A: u kno the box on the right hand side the 1 that has no neighbours” or “A: u know the L shape is the first box of that.” Path descriptions are similarly sensitive to the configuration of each maze and often use a landmark to identify where to start from but are distinguished by also, using a move by move route that can be followed through the particular configuration of boxes and links to a target location, for example, “A: go left twice and down twice.”

Line and Matrix descriptions, by contrast, have a more systematic organization. In Line descriptions this involves an ordered row or column elements with target locations identified as boxes inside these elements. For example “A: my 1st switch is on the 2nd row, 2nd from the right” or “A: my x is at the 7th column and bottom square.” In the case of Matrix descriptions, the maze is treated as an array of boxes with locations specified, using Cartesian coordinates, for example, “A: am on 2,7, dest is 4,3.” Both these schemes involve systematic enumeration of elements with ordinal counting in the case of Line descriptions and cardinal in the case of Matrix.

#### 4.1 Semantic coordination in the Maze Task

In addition to differences in their form, Healey (2008) proposes that the different maze description schemes are also associated with different semantic models. The Figural descriptions involve a relatively ad hoc ontology of salient landmarks or groupings that



does not systematically generalize across mazes. Path descriptions also depend on the specific configuration of a particular maze. For example, “gaps” are not used as part of a path, although the concept of a stepwise pathway does generalize across mazes. The limited systematicity of these two description types is illustrated by way of what is ostensibly the same location on two different mazes rarely receives the same Figural or Path description. In contrast to this, Line and Matrix description schemes involve more abstract ontologies of possible locations. In a Line description, locations are organized into an ordered set of rows or columns which can be applied in the same way to all instances of the maze. In the case of Matrix descriptions, the ontology is even more uniform with locations organized as a simple grid-like array with two axes. These more systematic structures generalize to all instances of the maze and are sufficiently abstract that missing boxes or gaps in the maze are still enumerated when a description is produced. In this case, the same location on any two mazes is normally associated with the same form of Line or Matrix description.<sup>1</sup> Healey (2008) draws on these observations to propose a basic semantic ordering on the description types: Figural, Path, Line, Matrix, which goes from the relatively concrete, instance-specific forms to the more abstract, systematic, and generalizable forms.

#### 4.2 Coordinating language use in the Maze Task

The standard pattern of coordination observed in the Maze Task is that people migrate from using Figural and Path descriptions in early trials toward more abstract Line and Matrix descriptions in later trials (Garrod & Anderson, 1987). This pattern is illustrated by the sequence of descriptions in Table 1.

What drives this shift in referring expressions over time? Previous work has shown that it is not, in general, achieved by explicit negotiation; for example not by stating that the maze is a grid with numbers down the side and letters across the bottom (Garrod & Anderson, 1987; Garrod & Doherty, 1994; Healey, 1997). People do sometimes attempt to coordinate using explicit negotiation, but it usually fails and people often violate whatever approach they have just explicitly agreed on. In general, this approach appears to be

Table 1  
Example evolution of Maze Task referring expressions

0 min	The sticking out piece on top of the maze
2 min	Left hand corner of the maze
5 min	The northernmost box
10 min	Topmost square of the column on top
15 min	3rd column middle square
20 min	3rd column first square
25 min	6th row longest column
30 min	6th row 1st column
40 min	6 r 1 c
45 min	6,1



of most use once people have already coordinated on a basic semantic model for the maze and is more commonly observed in later trials on the task where people have already built up some coordination (Healey, 1997).

In lieu of explicit negotiation, explanations of coordination in the Maze Task, such as Input-Output Coordination (Garrod & Anderson, 1987) and Interactive Alignment (Pickering & Garrod, 2004, 2007), have proposed Aggregate mechanisms of coordination. At any particular point in the conversation, the activation of representations involved in comprehending a description of a given type makes the production of a new description of the same type much more likely. So if one person produces, say, a Path description of a location, this will also activate the lexical, syntactic and semantic representations associated with those descriptions for the addressee who hears it. When they subsequently produce a description of another location, these representations are already active and this facilitates the production of a new Path style description. In this way, language processing becomes entrained across participants.

A problem for this kind of explanation is that it doesn't predict the pattern of migration captured in Table 1. People initially have most success in this task with the Figural and Path descriptions and statistically these will be most strongly activated; nevertheless, as noted, people consistently migrate away from these toward Line and Matrix description types (Garrod & Anderson, 1987; Garrod & Doherty, 1994; Healey, 1997). This trend is counter to what is predicted by priming, precedent, or recency effects alone. A second difficulty is that this picture of coordination is too passive. At the start of the Maze Task people often differ on the meaning of even relatively simple words like "box" or "row" in the context of the maze. Healey (1997) found that around 65% of people's initial referring expressions in the Maze Task are subject to some form of explicit clarification request. These overt, active clarification and repair processes seem, by definition, to require an Interactional explanation. A third issue is that Aggregate accounts presuppose that the lexical, syntactic, and semantic resources needed to produce or comprehend each scheme are in some sense already available to participants. The coordination problem is modeled primarily as a matter of ensuring the same scheme is activated. However, there is evidence that the solutions to the coordination problems which people develop in this task are, to a significant degree, novel conventions that emerge from the interaction. For example, there is evidence that novel and semantically distinct task sub-languages emerge by default in different groups who perform the task separately (Healey, 1997, 2008).

### *4.3 Hypotheses*

To summarize, the Running Repairs Hypothesis combines the claim that interactive feedback is central to the coordination of language use in conversation with the claim that negative feedback plays an especially critical role in coordinated language use by underpinning coordinated semantic change. This leads to two basic experimental hypotheses about the patterns of language use and language change observed in the Maze Task:

H1: Positive evidence of understanding promotes re-use of existing maze referring schemes.

H2: Negative evidence of understanding promotes evolution of new maze referring schemes.

We test these two hypotheses by selectively manipulating the kinds of feedback people appear to receive from each other. We use two manipulations which we label “Attenuation” and “Amplification.” The first tests H1 by reducing the degree of positive evidence a person receives in response to a location description. The second tests H2 by increasing the strength of the negative evidence someone receives by replacing a relatively precise clarification question with a less specific or “stronger” repair initiation that signals a potentially wider range of possible problems (Drew, 1997; Sacks, Schegloff, & Jefferson, 1974).

## 5. Methods

The manipulations of live Maze Task dialogues were carried out using a customized text chat tool (see Eshghi & Healey, 2016; Healey & Mills, 2006; Healey, Purver, King, Ginzburg, & Mills, 2003; Howes, Purver, Healey, Mills, & Gregoromichelaki, 2011) which enables substitutions of words in real-time interactions without participant’s knowledge. The two experimental manipulations are run between-subjects with a third control condition with no manipulations for comparison.

The use of text-chat enables experimental manipulations that are not possible with live spoken interaction. However, this also raises questions about whether results with text-chat generalize to other modalities. A full answer to this question requires comparison of specific interactional phenomena in each case (an issue taken up the discussion below), but the initial plausibility of this approach derives from the observation that many basic conversational mechanisms, such as clarification questions and self-repairs, do naturally occur in text-based interactions. Given the increasing importance of text messaging as a communication channel, even results that do not generalize beyond text-chat are significant for theories of human communication.

### 5.1 Materials

#### 5.1.1 The DiET chat tool

The experiments were conducted using a version of the Maze Task developed for the DiET chat tool experimental platform. As each turn is typed, the DiET server monitors it for target expressions, substitutes them with experimental probe expressions, and then sends the modified turn to the other participant. Only the recipient of the turn sees the substitution; the person who originally typed it does not. The list of target expressions and substitutions is generated from previous text-chat experiments to ensure natural text-chat usage is accommodated. This enables two experimental manipulations:

- 1. *Attenuation:* Identify naturally produced positive feedback such as “Yes” and “okay” and substitute with more provisional forms such as “erm,” “err,” “uurrm,” while the rest of the turn is left unchanged. For example, an “Okay” becomes “Err” and “Okay go to your goal now” becomes “Errr go to your goal now.”
- 2. *Amplification:* Identify naturally occurring reprise clarification requests that use a verbatim repeat of a word or phrase from a partner’s preceding turn followed by a question mark. For example, if A produces two successive turns: “A: I have two switches—one of them is on the top left” and “A: the other switch is in the top right” B might produce “top left?”, “top right?” or “top?” as reprise clarifications. These are substituted for less specific open class clarifications such as “What?”, “Huh?” and “Sorry?”

These manipulations preserve the basic structure of the turn sequence and leave the content of location descriptions unchanged while manipulating the evidence of understanding people appear to provide to each other. One difference between spoken and text conversation is that spoken questions like “What?” and “Sorry?” can be understood as signaling that somebody did not hear what was said. This interpretation is less likely in text-chat because turns persist.

Illustrative examples of these two manipulations drawn from the text chat corpus collected in the experiments reported below are provided in Tables 2, 3, 4, and 5. These excerpts show how participant’s interpretation of the manipulations depends on the immediate conversational context.

5.2 Predictions

The DiET chat tool makes a variety of potential dependent variables available for analysis. We use three simple measures of task performance (c.f. Healey, 2008; Howes et al., 2011). Number of Mazes completed within a fixed time (in this case 50 min) gives a

Table 2  
Example Attenuation 1, Path description (manipulation in bold, substituted words in ‘[]’)

A	Do u know where my position is?
B	<b>[yes]</b> → <b>uuumm</b>
A	I know where u are
B	U can only go left 2 or down left 1 down 1 ?
A	Left 1 down 1
B	Yes

Table 3  
Example Attenuation 2, Figural description (manipulation in bold, substituted words in ‘[]’)

A	Then im on the last box
B	On the far right hand side?
A	<b>[ye]</b> → <b>errrm</b>
B	All the way to the right i mean

Table 4

Example Amplification 1, Figural description (manipulation in bold, substituted words in '[ ]')

A	Ok i got it
A	Am on it now
B	<b>[ur on the top right?] → Sorry?</b>
A	On the shaded one
B	Ok where is ur shaded one describe it
B	Describe both quickly
A	Its on the top
B	Ok and the other one?...
A	And one after the gate
B	<b>BRO WE DONT HAVE THE SAME GATES DIDNT U LISTNE?</b>

Table 5

Example Amplification 2, Line description (manipulation in bold, substituted words in '[ ]')

A	Then where is your switch
B	Come to 2nd one
B	and tell me
A	<b>[2nd one what] → srry what?</b>
B	I mean come to 2nd block of last column
B	and stop there
A	No its blocked here

basic measure of global task performance. This is combined with two measures of task process. First, total number of turns used per maze completed, with turn completion operationalized as the text sent on pressing the *enter* key. This gives a simple measure of how much interaction is required to complete each maze. Second, typing speed in characters (including spaces) per second between onset of the first character and pressing *enter*. This provides a measure of how easy participants find it to formulate their turns, the assumption being that people who are confident of how to describe their position or what to do next will, all things being equal, type faster than those who are not (this approach is also used in Howes et al., 2011).

For measures of semantic coordination we use three dependent variables to assess the forms of referring expression used in turns that contain a location description. Following the previous Maze Task literature, we use the content based hand coding of description types: Figural, Path, Line, Matrix, following the classification described by Garrod and Anderson (1987) (see description above). In addition to this we provide two automatically coded measures. First, number of digits used in each location description that counts all instances of < 1st, 2nd, 3rd, 4th,... > and < 1, 2, 3, 4,... > in each turn (defined as above) but ignores number words such as “one,” “two,” “three,” “four,” and ordinals such as “first” and “last.” The rationale is that this provides a simple but fine-grained index of participants’ assumptions about how well coordinated their ontology of locations is. Roughly, if participants know what things to count and are doing it frequently enough to use digits, this suggests a relatively stable concept of what a “possible location” is in the maze and a relatively well-organized scheme (like the Line and Matrix schemes) for

individuating them. The last measure is description length measured as the number of characters (including spaces) in a turn containing a location description.

For example, “in the penultimate square on the left wing” is classified as a Figural description that uses 0 digits and 42 characters; “1st col 2nd row” is classified as a Line description containing 2 digits and 15 characters and “5,5” is a Matrix description using 2 digits and 3 characters.

Both interventions disrupt the dialogue and should therefore interfere with measures of task performance predicting fewer mazes completed and more turns to complete each maze. If the interventions make people unsure of what to say, this should, all things being equal, also lead to people taking more time to construct their turns.

The most important predictions relate to use of referring expressions. Different referring expressions have different forms, different typical lengths, and also make different use of enumeration (see illustration in Table 1). Enumeration is especially important because it indexes degree of coordination on the basic semantic ontology needed to underpin the more systematic semantic description schemes used by the Line and Matrix descriptions (see above).

H1 predicts that Attenuation should slow down the shift in description types, slow the increase in use of enumeration, and promote longer description types relative to controls. H2 predicts that Amplification should accelerate the shift in description types, increase enumeration, and shorten description types relative to controls.

### *5.3 Participants*

Participants were recruited anonymously from undergraduate and postgraduate students across a variety of disciplines using general mailing lists. A total of 88 participants were recruited in three separate rounds. They were divided into 18 pairs in the control condition, 13 pairs in the Amplification condition, and 13 in the Attenuation condition.

### *5.4 Procedure*

Pairs of subjects were booked in for 1 h slots. They were seated in separate experimental cubicles and given written instructions explaining the task. Each pair was asked to complete 12 pairs of mazes as quickly as possible. In the Attenuation condition three pairs of participants were allowed to continue for longer than in the other two conditions. To ensure comparability for the analysis we use only the data collected up to the first 50 min of each session. The experimental manipulations were not disclosed to the participants until the debriefing session.

## **6. Results**

The number of turns or mazes completed is not independent for the members of a pair. To allow for this the relevant data is analysed as scores per pair. For the rest of

the analyses pair is included as a random factor in order to allow for possible statistical dependencies between responses by the different members of a pair. Following Healey and Mills (2006); Eshghi and Healey (2016) we pool the scores for the first six *Early* and last six *Late* mazes to provide an index of task experience. This simplifies the analysis and also reduces the variability contributed by individual mazes. To accommodate the mix of repeated and between subjects measures and the different distributions of the dependent variables, Generalized Linear Mixed Models procedure in SPSS (v.21) was used for all statistical analyses other than simple non-parametric comparisons.

All spoof turns produced by the DiET server are excluded from all analyses to avoid biasing any of the typing measures; however, the original, unmodified turns are included in turn counts. One pair from the amplification condition is also excluded as the transcript indicated they had failed to understand the task. After these exclusions the three conditions together yielded a total corpus of 13,744 turns. Using the criteria described above, all turns were coded by the authors for whether they contained a location description (Y/N) and the type of location description (Figural, Path, Line or Matrix). This coding yielded a total of 2,844 location descriptions.

## 6.1 Task performance

Table 6 shows the average number of turns and mazes completed by each pair in the first 50 min of the interactions.

### 6.1.1 Turns

Mann–Whitney’s *U* test shows no significant difference between the Attenuation and Control conditions in number of turns taken per maze ( $U = -1.15$ ,  $p = .25$ ,  $N = 552$ ) nor between the Amplify and Control conditions ( $U = -0.76$ ,  $p = .45$ ,  $N = 557$ ).

### 6.1.2 Mazes

Chi squared comparisons of the number of pairs who did or did not reach the last Maze of the 12 within 50 min also showed no significant differences (Attenuation vs. Control =  $Chi(1)^2 = 1.41$ ,  $p = .24$ ; Amplification vs. Control  $Chi(1)^2 = 0.54$ ,  $p = .46$ ). However, because this is a measure of the pair’s performance the numbers in each cell are small and the comparisons are low powered.

Table 6  
Summary statistics

Averages	Control	Attenuate	Amplify
Average number of turns per maze	34.2	37.4	34.2
Average number of mazes completed	9.5	7.2	8.2

6.1.3 Typing speed

Inspection of the data showed some turns had typing speeds above 7 characters per second (the level achieved by professional typists) these had been created by participants occasionally keeping a key continuously pressed. To filter these out, all turns with speeds over 7 characters per second were excluded from the analysis. Participants’ average typing speed on each maze were analyzed in two GLMM Linear analyses with Pair as a random factor and Stage (First 6 vs. Last 6), Condition (Attenuation or Amplification vs. Control) and the Stage  $\times$  Condition interaction as fixed factors.

For Attenuation, there is no main effect of Condition ( $F_{(1,534)} = 2.32, p = .13$ ), a main effect of Stage ( $F_{(1,534)} = 38.6, p < .01$ ), and no Condition  $\times$  Stage Interaction ( $F_{(1,534)} = 0.03, p = .87$ ). For Amplification there is also no main effect of Condition ( $F_{(1,538)} = 0.22, p = .64$ ), a main effect of Stage ( $F_{(1,538)} = 16.82, p < .01$ ), and no Condition  $\times$  Stage Interaction ( $F_{(1,538)} = 0.04, p = .85$ ).

Table 7 illustrates the overall pattern for typing speed. Typing speed for all participants increases as the task progresses. Although the means suggest some impact of Attenuation on typing speed, the effect is not statistically significant.

6.2. Language use

6.2.1 Description types

Comparison of the distribution of description types across conditions was made with two multinomial regression analyses with Pair as a random factor and Stage (First 6 vs. Last 6) and Condition (Attenuation or Amplification vs. Control) and the Stage  $\times$  Condition interaction as fixed factors. Multinomial regression estimates the relative probabilities of the distribution of description types and so differences in the total number of descriptions at different stages do not affect the analysis.

Attenuation shows no main effect of Condition ( $F_{(4,108)} = 1.13, p = .34$ ), a main effect of Stage ( $F_{(4,108)} = 12.86, p < .01$ ) and no Condition  $\times$  Stage Interaction ( $F_{(4,108)} = 0.45, p = .77$ ). Amplification shows both a main effect of Condition ( $F_{(4,108)} = 8.75, p < .01$ ) and a main effect of Stage ( $F_{(4,108)} = 20.3, p < .01$ ) and no Condition  $\times$  Stage Interaction ( $F_{(4,108)} = 0.45, p = .77$ ).

Fig. 2 illustrates the pattern of results. The Control and Attenuation conditions follow the typical pattern of migration from Figural and Path descriptions toward Line and Matrix (see above) and are not reliably different from each other. Pairs in the

Table 7  
Average typing speeds (Estimated Marginal Means)

Average Characters per Second	Stage 1	Stage 2
Control	2.99	3.17
Attenuate	2.55	2.72
Amplify	2.85	3.01



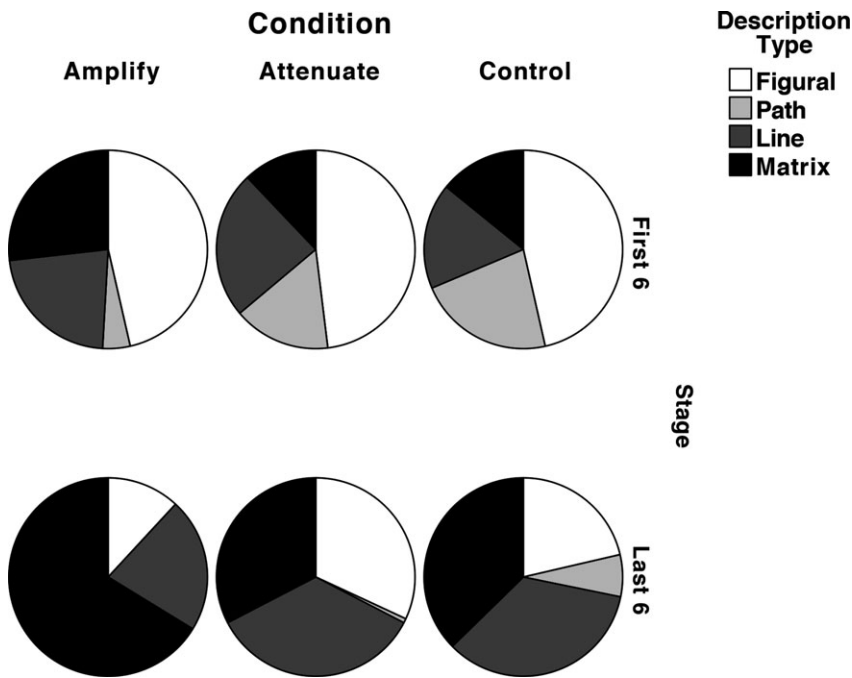


Fig. 2. Pattern of use of location description types.

Amplification condition make the shift to abstract description types more strongly and more quickly.

The counts of digits were positively skewed so a GLMM model with a Poisson distribution was used to compare the frequency of digits in location descriptions across conditions. As above, pair was included as a random factor, Stage and Condition (Attenuation or Amplification vs. Control) and the Stage  $\times$  Condition interaction as fixed factors.

Attenuation shows no main effect of Condition ( $F_{(1,2120)} = 0.03$ ,  $p = .87$ ); a main effect of Stage ( $F_{(1,2120)} = 92.2$ ,  $p < .01$ ); and no interaction ( $F_{(1,2120)} = 0.94$ ,  $p = .33$ ). Amplification also shows no main effect of Condition ( $F_{(1,2013)} = 0.03$ ,  $p = .23$ ); a main effect of Stage ( $F_{(1,2013)} = 102$ ,  $p < .01$ ); and in addition, a Stage  $\times$  Condition interaction ( $F_{(1,2013)} = 5.75$ ,  $p = .02$ ). The source of the interaction is pairs in the Amplification condition, using reliably more digits in the last six mazes (Fixed Coefficient = 0.24,  $t = 2.413$ ,  $p = .02$ ). The overall pattern of digit use is illustrated in Fig. 3.

## 7. Discussion

The experimental manipulations do not have the predicted effects on task performance. Although the means are suggestive, the number of mazes completed, the number of turns

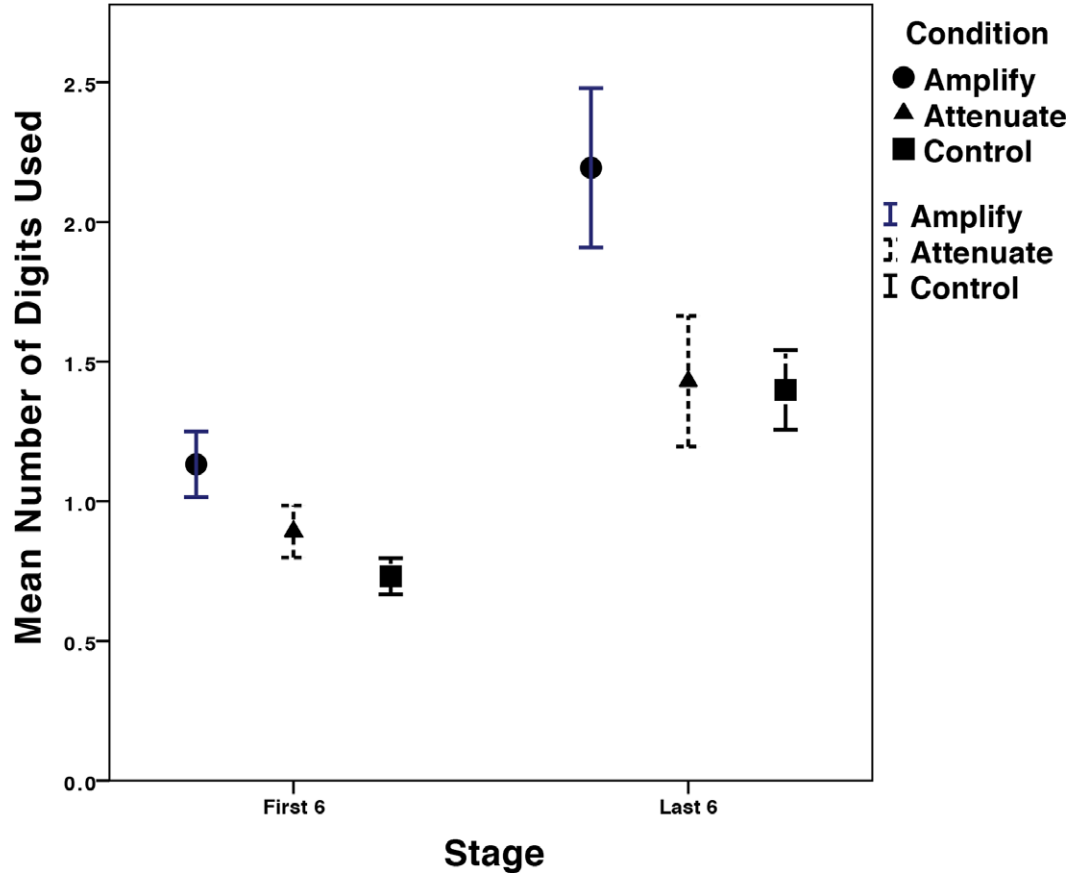


Fig. 3. Use of enumeration in location descriptions (Error Bars show 95% C.I.).

used per maze, and the speed with which people construct their turns are not reliably different across conditions. In general, we expect effects on task performance because differences in people’s degree of linguistic coordination should ultimately translate to differences in task performance. One possibility is that this assumption doesn’t hold for this task. People might be able to use actions in the maze to compensate for difficulties with communication—for example, using moves in the maze to test their interpretation of a location description rather than questions. Another possibility is that 12 trials provide insufficient statistical power for reliable effects on performance to emerge. A longer sequence of trials and possibly a larger number of participants would be needed to test this possibility.

The effects on semantic coordination are more interesting. The reduction in positive evidence does not reliably affect choice of description type. This failure to find an effect does not directly falsify or support the hypothesis that positive evidence of understanding promotes re-use of a referring expression (H1). It does, however, suggest that positive feedback is not a critical mechanism driving changes in choice of location description types.

This result appears to conflict with previous findings that patterns of change in definite referring expressions are directly affected by the availability and strength of positive feedback (Clark & Schaefer, 1989; Schober & Clark, 1989; Clark & Wilkes-Gibbs, 1986; Metzger & Brennan, 2003). The simplest potential explanation for this is that the experimental intervention used here is not strong enough to uncover these effects, or there was insufficient statistical power. The examples in Table 2 and 3 suggest that the Attenuation intervention did not have a marked effect on the turn sequence.

A second possible explanation for the conflict with previous findings is the difference between the Maze Task and the typical definite reference tasks. In tasks such as the Tangram Task and its variants, people are required to coordinate repeated references to the same specific items. In contrast to this, the Maze Task requires people to coordinate references to a succession of different specific instances of items of a similar type, sequences of locations. It is possible that positive evidence of understanding may be more important to the collaborative processes of establishing a name for a recurring item, that is, for the processes of contraction or abbreviation typically seen in the Tangram Task (Clark & Schaefer, 1989; Clark & Wilkes-Gibbs, 1986), than it is for developing general referring schemes that work across non-repeating instances of a class of things that is, the process of abstraction seen in the Maze Task (Garrod & Anderson, 1987; Healey, 2008).

In contrast to this, the Amplification manipulation causes a strong and early shift to more abstract description types. This supports the hypothesis that negative evidence plays an important role in coordinating the evolution of maze referring schemes (H2). It shows, in particular, that the effect on description types is mediated by differences between the repair sequences that the original clarification questions would, counter-factually, have prompted and the sequences that the experimental substitutions actually caused. Convergence on more systematic and abstract semantic models of locations is facilitated by, in effect, highlighting misunderstandings.

This finding also raises a question about the Running Repairs Hypothesis, which is built on the intuition that negative evidence is useful because it provides localized, public information about likely differences in language use that can be used to help people adapt in response to problems with mutual intelligibility (Healey, 1997, 2008; Mills, 2013). However, the Amplification manipulation used here strictly reduces the information available to participants; substitution of relatively specific clarification questions with “unrestricted” repair initiations like “what?” and “sorry?” should make it harder for an addressee to localize or diagnose what aspects of a prior turn their partner has a problem with (Sacks et al., 1974b). This is puzzling but our general proposal is that the effect of the amplifications is to take a real problem encountered in the dialogue and then prompt the participants to work harder to resolve it than they otherwise would. Some indication of this can be seen in the excerpts in Tables 4 and 5 where the trajectory of the turn sequence does appear to be substantially altered by the intervention. Although we cannot be sure that they would not have happened anyway, it seems plausible that the elaborations “on the shaded one” and “i mean come to 2nd block of last column” are prompted by the intervention and that answers to the actual, substituted questions would have been

different. These responses each provide useful additional information that would help participants discover and hunt down possible differences in interpretation and, in doing so, converge more rapidly (cf. Brennan & Schober, 2001).

These results establish a *prima facie* case for a direct causal connection between local repair processes and the evolution of semantic conventions over the course of a conversation. To make this general line of explanation work requires, among other things, the development of semantic models that can model such on-the-fly conceptual adaptations on a turn-by-turn, and even word-by-word basis to capture the proposed process of adjustment of interpretations. Although such a model does not yet exist, advances in formal semantics suggest it is not out of reach (see, for example, Ginzburg & Kolliaku, this issue; Larsson, this issue; also Eshghi et al. [2015]; Howes and Eshghi [2017] for a semantic model of feedback that is compatible with the Running Repairs Hypothesis).

The results also suggest the potential for a productive new interface between Conversation Analysis, Formal Semantics, and Psycholinguistics. The detailed structure of repair sequences is well understood (Schegloff, 1987, 1992; Schegloff, Jefferson, & Sacks, 1977). However, their effects on semantic change have not previously been explored experimentally. The fine-grained, real-time interventions made possible by the DiET chat tool enable selective interference with different repair processes. This enables a form of experimental conversation analysis that, we hope, can complement existing work by causally testing the effects of a variety of different interactional mechanisms on the subsequent trajectory of an interaction.

Some caveats are needed. One methodological issue with this approach is the use of text-based interaction. Text chat in its various forms is now a common form of communication, but we know relatively little about how processes such as turn-taking and repair in text-chat formats differ from spoken conversation (Schoönfeldt & Golato, 2003).<sup>2</sup> The plausibility of the generalization made here is supported by the conversational style of the Maze Task chat exchanges, including the natural production of simple grounding signals and repair phenomena on which the experimental manipulations depend. It is also supported by the similarity in patterns of maze description type produced in text-chat and spoken versions of the task (see, e.g., Garrod & Anderson, 1987; Garrod & Doherty, 1994; Healey, 1997). Nonetheless, only further work can determine whether this is a serious problem.

The ideal response would be to extend the techniques used by DiET to spoken interaction to test whether the same results are obtained. However, there is a practical difficulty. The experiments require that what has been said is recognized before an intervention is triggered. For example, the spoken target “Okay” is recognized before substituting it for a spoof “Ummm.” This introduces a delay that significantly disrupts the dynamics of live spoken or non-verbal interaction but is compatible with the feedback cycles typical of text-based interaction.

Another general methodological concern is the use of task-oriented dialogues instead of the naturally occurring conversation data favored by conversation analysis. This is not an intrinsic limitation of the DiET method which can equally well be used in unstructured text interactions. The particular value of using task-oriented dialogues here is that it

provides a relatively well-understood, independent measure of semantic coordination. Comparable measures are not currently available for natural dialogue.

This leads to the critical question of whether the present findings generalize to other tasks and situations. The Running Repairs Hypothesis makes a general claim about the primary mechanisms that coordinate language processing in dialogue and it is a large step to generalize from the present experimental results to processes of language coordination in general. Some convergent support comes from prior work that shows that patterns of (spoken) repair are closely correlated with semantic coordination in the Maze Task (Healey, 1997) and that interfering with the resources available for identifying and resolving communication problems systematically alters semantic coordination in both a spoken version of the Maze task and in graphically mediated interactions (Healey, 2008; Healey, Swoboda, Umata, & King (2007)).

There is also good case for testing these ideas in a wider range of tasks and situations. The mechanisms of clarification and repair appear to be a ubiquitous and universal feature of natural dialogue (Colman & Healey, 2011; Dingemanse et al., 2013; Kendrick, 2015; Schegloff, 2007) and for this reason alone should attract much more attention in experimental and theoretical psycholinguistics. This study provides an additional reason for giving repair a primary role in explaining coordinated language processing in dialogue.

If misunderstandings and the repairs that follow them are an important driver of semantic change, this has a significant corollary: It makes this form of language change a constitutively joint, interactional process. Sequences of clarification and repair only happen when different people's language processing capabilities come into contact and mismatches between them are exposed. This also entail that the local patterns of language change that result from these joint processes are specific to the particular people and the particular difficulties they actually encounter in the process of coordinating on a particular task. If correct, this points to a limitation in "globalist" approaches to modeling language, used in Machine Learning and other forms of statistical modeling which idealize away from individual differences and the adaptive processes people use in dealing with them. These local processes may well turn out to be a key advantage of natural languages by enabling flexible and creative adaptation to new people, new situations, and new tasks.

## 8. Conclusions

Positive and negative evidence of understanding appear to play systematically different roles in people's ability to converge on particular description schemes in the Maze Task. In particular, while Attenuation of positive evidence has no significant effect, Amplification of negative evidence promotes convergence on systematic, abstract ways of describing maze locations. The processes involved in detecting and repairing misunderstandings do more than keep the conversation afloat (Schegloff, 1992), they have systematic effects on subsequent language use by driving changes in the speed and form of semantic coordination that emerges. This provides support for the Running Repairs Hypothesis: Coordination of

language use depends first and foremost on processes used to deal with misunderstanding on the fly. The implication is that language processing in dialogue involves mechanisms that are qualitatively different from those needed for individual language processing. This also suggests an opportunity for developing new conceptual and empirical connections between Conversation Analysis, Psycholinguistics, and Formal Semantics.

## Acknowledgments

This research was partly supported by ESRC through the Dynamics of Conversational Dialogue project (RES-062-23-0962) and EPSRC through the Dialogue Experimentation Toolkit (DiET) project (EP/D057426/1). Eshghi has also been gratefully supported by the EPSRC, under grant number EP/M01553X/1 (BABBLE project).

## Notes

1. It is important to note that within each description type as described here, there is significant fine-grained variation. For example, in Matrix schemes, there is variation in the choice of origin (0 or 1) which sides form the axes, use of letters, or numbers for each axis, etc.
2. Though see Clark and Brennan (1991) for an early discussion of how different media place different constraints on grounding and repair.

## References

- Anderson, A., & Garrod, S. (1987). The dynamics of referential meaning in spontaneous dialogue. In R. G. Reilly (Ed.), *Communication failure in dialogue and discourse* (pp. 161–183). Amsterdam: Elsevier.
- Barr, D. J., & Keysar, B. (2002). Anchoring comprehension in linguistic precedents. *Journal of Memory and Language*, 46, 391–418.
- Bavelas, J. (2007). *Face-to-face dialogue as a micro-social context. the example of motor mimicry*. Gesture and the dynamic dimension of language. Essays in honor of David McNeill, vol. 1. Amsterdam: John Benjamin, 127–146.
- Bavelas, J. B., Coates, L., Johnson, T. (2000). Listeners as co-narrators. *Journal of Personality and Social Psychology*, 79(6), 941–952.
- Brennan, S., & Schober, M. (2001). How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Language*, 44(2), 274–296.
- Christiansen, M., & Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, 39, e62.
- Clark, H. H. (1996). *Using language*. Cambridge, UK: Cambridge University Press.
- Clark, H. H. (1998). Communal lexicons. In K. Malmkjaer & J. J. Williams (Eds.), *Context in language learning and language understanding* (pp. 63–87). Cambridge, UK: Cambridge University Press.
- Clark, H. H., & Brennan, S. A. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). Washington: APA Books.

- Clark, H. H., & Schaefer, E. A. (1989). Contributing to discourse. *Cognitive Science*, 13, 259–294.
- Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22, 1–39.
- Colman, M., & Healey, P. G. T. (2011). The distribution of repair in dialogue. In T. Verhoef et al. (Eds.), *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society* (pp. 1563–1568). Boston, MA: Cognitive Science Society.
- Dingemanse, M., Torreira, F., & Enfield, N. J. (2013). Is huh? a universal word? Conversational infrastructure and the convergent evolution of linguistic items. *PLoS ONE*, 8(11), e78273.
- Drew, P. (1997). Openclass repair initiators in response to sequential sources of troubles in conversation. *Journal of Pragmatics*, 28(1), 69–101.
- Eshghi, A., & Healey, P. G. T. (2016). Collective contexts in conversation: Grounding by proxy. *Cognitive Science*, 40, 299–324.
- Eshghi, A., Howes, C., Hough, J., Gregoromichelaki, E., & Purver, M. (2015). Feed-back in conversation as incremental semantic update. In M. Purver, M. Sadrzadeh & M. Stone (Eds.), *Proceedings of the 11th International Conference on Computational Semantics (IWCS)* (pp. 261–271). Association for Computational Linguistics.
- Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, 27, 181–218.
- Garrod, S., & Doherty, G. (1994). Conversation, co-ordination and convention: An empirical investigation of how groups establish linguistic conventions. *Cognition*, 53, 181–215.
- Ginzburg, J. (2012). *The interactive stance: Meaning for conversation*. Oxford, UK: Oxford University Press.
- Ginzburg, J., & Kolliaku, D. (2018). Divergently seeking clarification: The emergence of clarification interaction. *Topics in Cognitive Science*, 10(2), 335–366.
- Goodwin, C. (1979). The interactive construction of a sentence in natural conversation. In G. Psathas (Ed.), *Everyday language: Studies in ethnomethodology* (pp. 97–121). New York: Irvington.
- Healey, P. G. T. (1997). Expertise or expertese: The emergence of task-oriented sub-languages. In M. Shafto & P. Langley (Eds.), *Proceedings of the 19th Annual Conference of the Cognitive Science Society* (pp. 301–306). Stanford, CA: Stanford University.
- Healey, P. G. T. (2008). Interactive misalignment: The role of repair in the development of group sub-languages. In R. Cooper & R. Kempson (Eds.), *Language in flux*. College Publications.
- Healey, P. G. T., Colman, M., & Thirlwell, M. (2005). Analysing multi-modal communication: Repair-based measures of human communicative co-ordination. In J. van Kuppevelt, L. Dybkjaer, & N. Bernsen (Eds.), *Natural, intelligent and effective interaction in multimodal dialogue systems*. Vol. 30 (pp. 113–129), *Text, Speech and Language Technology*. Dordrecht, the Netherlands: Kluwer.
- Healey, P. G. T., & Mills, G. J. (2006). Participation, precedence and co-ordination in dialogue. In R. Sun (Ed.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 1470–1475). Mahwah, NJ: Erlbaum.
- Healey, P. G. T., Purver, M., King, J., Ginzburg, J., & Mills, G. J. (2003). Experimenting with clarification in dialogue. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*, Boston, MA: Erlbaum.
- Healey, P. G., Swoboda, N., Umata, I., & King, J. (2007). Graphical language games: Interactional constraints on representational form. *Cognitive Science*, 31(2), 285–309.
- Horton, W., & Gerrig, R. (2005). Conversational common ground and memory processes in language production. *Discourse Processes*, 40(1), 1–35.
- Howes, C., & Eshghi, A. (2017). Feedback relevance spaces: The organisation of increments in conversation. In C. Gardent & C. Retore (Eds.), *Proceedings of the 12th International Conference on Computational Semantics (IWCS)*.
- Howes, C., Purver, M., Healey, P. G. T., Mills, G. J., & Gregoromichelaki, E. (2011). On incrementality in dialogue: Evidence from compound contributions. *Dialogue and Discourse*, 2(1), 279–311.
- Kendrick, K. H. (2015). Other-initiated repair in english. *Open Linguistics*, 1(1), 164–190.



- Krauss, R. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of personality and social psychology*, 4(3), 343.
- Larsson, S. (2018). Grounding as a side-effect of grounding. *Topics in Cognitive Science*, 10(2), 389–408.
- Lewis, D. (1969). *Convention*. Cambridge, MA: Harvard University Press.
- Metzing, C., & Brennan, S. (2003). When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language*, 49(2), 201–213.
- Mills, G. J. (2013). Dialogue in joint activity: Complementarity, convergence and con-ventionalization. *New Ideas in Psychology*, 32, 158–173.
- Pickering, M., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27, 169–226.
- Pickering, M., & Garrod, S. (2006). Alignment as the basis for successful communication. *Research on Language and Computation*, 4, 203–228.
- Pickering, M., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, 11(3), 105–110.
- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36, 329–347.
- Purver, M., Ginzburg, J., & Healey, P. G. T. (2003). On the means for clarification in dialogue. In R. Smith & van Kuppevelt, J. (Eds.), *Current and new directions in discourse & dialogue* (pp. 235–255). Dordrecht, Netherlands: Kluwer Academic.
- Sacks, H., Schegloff, E., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4), 696–735.
- Schegloff, E. (1987). Some sources of misunderstanding in talk-in-interaction. *Linguistics*, 25, 201–218.
- Schegloff, E. (1992). Repair after next turn: The last structurally provided defense of intersubjectivity in conversation. *American Journal of Sociology*, 97(5), 1295–1345.
- Schegloff, E. A. (1993). Reflections on quantification in the study of conversation. *Research on Language and Social Interaction*, 26(1), 99–128.
- Schegloff, E. (1995). Parties and talking together: Two ways in which numbers are significant for talk-in-interaction. *Situated Order: Studies in the Social Organization of Talk and Embodied Activities*, 31–42.
- Schegloff, E. A. (2007). *Sequence organization in interaction: Volume 1: A primer in conversation analysis*. Cambridge, UK: Cambridge University Press.
- Schegloff, E., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, 53(2), 361–382.
- Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive psychology*, 21(2), 211–232.
- Schoöfeldt, J., & Golato, A. (2003). Repair in chats: A conversation analytic approach. *Research on language and social interaction*, 36(3):241–284.
- Schober, M. F., & Brennan, S. E. (2003). Processes of interactive spoken discourse: The role of the partner. In A. C. Graesser, M. A. Gernsbacher, & S. R. Goldman (Eds.), *Handbook of discourse processes* (pp. 123–164). Hillsdale, NJ: Lawrence Erlbaum.
- Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of Memory and Language*, 31, 183–194.
- Yngve, V. H. (1970). On getting a word in edgewise. In M. A. Campbell (Ed.), *Papers from the 6th regional meeting of the Chicago Linguistic Society* (pp. 567–578). Chicago, IL: University of Chicago.